JOURNAL OF MATHEMATICAL ANALYSIS AND APPLICATIONS 47, 256–268 (1974)

"Envelope Programming" and Conjugate Duality

L. McLINDEN*

Mathematics Research Center, University of Wisconsin, Madison, Madison, Wisconsin 53706†

Submitted by George Leitmann

In a recent paper D. J. White presented a new approach to the problem of minimizing a differentiable convex function over a convex set. The idea begins with describing the convex function as the envelope of its tangent hyperplanes. With this description the given problem is represented in "min–max" form. An appeal to White's minimax theorem then permits one to interchange the extrema and arrive at a dual problem having "max–min" form. In the present paper White's approach is first generalized and analyzed and then related to well-known results in conjugate duality.

1. INTRODUCTION

Consider the problem of minimizing a convex function f over a convex subset D of a real topological vector space X, i.e.,

$$\min_{x \in D} f(x). \qquad (\mathscr{P})$$

For $X = R^n$ and f differentiable, D. J. White recently showed [9] that solving $(\mathscr{P})$ is equivalent to solving the saddle point problem

$$\operatorname*{minimax}_{x \in D\ u \in X} H(x, u),$$

where $H(x, u) = f(u) + \langle x - u, \nabla f(u)\rangle$. Since f is the upper envelope of the tangent hyperplanes to its graph, $f(x) = \sup_{u \in X} H(x, u)$, and hence $(\mathscr{P})$ corresponds to the upper half of the minimax problem, i.e., to

$$\min_{x \in D}\{\sup_{u \in X} H(x, u)\}.$$

On the other hand, the lower half of the minimax problem, i.e.,

$$\max_{u \in X}\{\inf_{x \in D} H(x, u)\}.$$

* Sponsored by the United States Army under Contract No. DA-31-124-ARO-D-462.

† Present address: University of Illinois, Urbana, IL 61801.

may be interpreted as a (nonconcave) maximization problem dual to ($\mathscr{P}$). The term "envelope programming" was introduced by White to describe this overall approach to ($\mathscr{P}$). The novelty of his approach resides in the choice of H as the function determining an equivalent saddle point problem.

This paper addresses three questions: (i) To what extent can White's saddle point characterization be generalized, (ii) how extensive a duality theory for ($\mathscr{P}$) can be developed, taking the function H (or a suitable generalization) as the "Lagrangian"; and (iii) in what way does envelope programming relate to existing duality theory?

The paper is in two parts. In Section 2 we deal with questions (i) and (ii). Concerning (i), we show that White's saddle point characterization extends to the case in which X has arbitrary dimension and f is not necessarily differentiable. For a nondifferentiable f this entails substituting the notion of supporting hyperplane for tangent hyperplane. Concerning (ii), examples show that there may be a finite or infinite "duality gap" even though ($\mathscr{P}$) satisfies strong constraint qualifications. Also mentioned are several other obstacles to a general duality theory built on the choice of H as the "Lagrangian" function. In Section 3 we deal with question (iii). There it is shown that envelope programming is fairly closely related to that body of results in conjugate duality which pertains to Fenchel's Duality Theorem.

2. Envelope Programming

Let X be a Hausdorff locally convex topological vector space over the real number system R, and let X^* be the (continuous) dual of X. For any $x \in X$ and $x^* \in X^*$, let $x^*(x)$ be denoted by $\langle x, x^*\rangle$. Let D be a nonempty convex subset of X, and let f be a proper convex function on X. Thus, f is an everywhere-defined function with values in $R \cup \{+\infty\}$, which is not identically $+\infty$, and which satisfies

$$f((1-\lambda)\,x + \lambda y) \leqq (1-\lambda)\,f(x) + \lambda f(y)$$

for all $\lambda \in [0, 1]$ and $x, y \in X$. Equivalently, the set $\{(x, \mu) \in X \times R \mid f(x) \leqq \mu\}$ lying on or above the graph of f is nonempty and convex. The "nonvertical" supporting (closed) hyperplanes to this set can be indexed by

$$\Gamma = \{(u, u^*) \in X \times X^* \mid f(x) \geqq f(u) + \langle x - u, u^*\rangle, \forall x \in X\}.$$

The multivalued mapping ∂f from X to X^* having graph Γ is called the *subdifferential* of f. It is known that for any point $x \in X$ at which f is differentiable (in the sense of Gâteaux), the image set $\partial f(x)$ under the subdifferential is precisely the singleton $\{\nabla f(x)\}$; conversely [3, Sect. 10g], if f is finite and continuous at a point x for which $\partial f(x) = \{x^*\}$, then f is differentiable at x with gradient x^*. In particular, then, if f is differentiable throughout X, then Γ is exactly the graph of the gradient mapping ∇f.

In order to extend White's saddle point characterization to the non-differentiable case, we introduce a function J, given by

$$J(x, u, u^*) = f(u) + \langle x - u, u^* \rangle,$$

to replace his function H and consider the saddle point problem

$$\operatorname*{minimax}_{x \in D\, (u,u^*) \in \Gamma} J(x, u, u^*). \qquad (\mathscr{L}_W)$$

By the comments above concerning Γ, this problem is equivalent to White's when f is differentiable. We shall say that $(\mathscr{P})$ *satisfies the constraint qualification* if and only if (a) there exists a point of D at which f is finite and continuous, or (b) there exists a point interior to D at which f is merely finite. It is helpful also to single out the (local) condition

$$f(x) = \sup_{(u,u^*) \in \Gamma} \{f(u) + \langle x - u, u^* \rangle\}, \qquad (*)$$

which we shall describe by saying f *satisfies* $(*)$ *at* x. Notice that f is the upper envelope of its supporting hyperplanes precisely when f satisfies $(*)$ everywhere, i.e., for every $x \in X$. We can now state the relationship between $(\mathscr{L}_W)$ and $(\mathscr{P})$.

THEOREM. *If $\bar{x}$ solves $(\mathscr{P})$ and $(\mathscr{P})$ satisfies the constraint qualification, then there exists a pair $(\bar{u}, \bar{u}^*)$ such that $(\bar{x}, \bar{u}, \bar{u}^*)$ solves $(\mathscr{L}_W)$ and the two optimal values are equal. Conversely, if $(\bar{x}, \bar{u}, \bar{u}^*)$ solves $(\mathscr{L}_W)$ and f satisfies $(*)$ at $\bar{x}$, then $\bar{x}$ solves $(\mathscr{P})$ and the two optimal values are equal.*

The theorem is almost immediate from the two lemmas which follow. The assertions concerning equality of optimal values follow from the lemmas and the fact that, when $(\bar{x}, \bar{u}^*) \in \Gamma$, the condition $(\bar{u}, \bar{u}^*) \in \Gamma$ is equivalent to $f(\bar{x}) = J(\bar{x}, \bar{u}, \bar{u}^*)$. The first lemma is well known (e.g. [7, Theorem 27.4] or [6, Corollary 1 to Theorem 1]).

LEMMA 1. *A sufficient condition for $\bar{x} \in D$ to solve $(\mathscr{P})$ is that there exists $\bar{u}^*$, such that $-\bar{u}^*$ is normal to D at $\bar{x}$ and $(\bar{x}, \bar{u}^*) \in \Gamma$. If $(\mathscr{P})$ satisfies the constraint qualification, this condition is also necessary.*

Proof. If $-\bar{u}^*$ is normal to D at $\bar{x}$, then

$$\langle x - \bar{x}, -\bar{u}^* \rangle \leqq 0, \qquad \forall x \in D,$$

while if $(\bar{x}, \bar{u}^*) \in \Gamma$, then

$$f(\bar{x}) \leqq f(x) + \langle x - \bar{x}, -\bar{u}^* \rangle, \qquad \forall x \in X.$$

Together, these inequalities imply that

$$f(\bar{x}) \leqq f(x), \qquad \forall x \in D.$$

Now assume $\bar{x}$ solves $(\mathscr{P})$, and define h to be the function which is 0 on D and $+\infty$ on $X \backslash D$. Then $\bar{x}$ yields an unconstrained minimum of $f + h$, so that $0 \in \partial(f + h)(\bar{x})$. If $(\mathscr{P})$ satisfies the constraint qualification, then the subdifferential formula $\partial(f + h) = \partial f + \partial h$ is valid ([4, Theorem 3(b)] or [3, Sect. 10d]), and hence $0 \in \partial f(\bar{x}) + \partial h(\bar{x})$. This implies $(\bar{x}, \bar{u}^*) \in \Gamma$ for some $\bar{u}^*$ such that $-\bar{u}^* \in \partial h(\bar{x})$. Since $\partial h(\bar{x})$ is the cone of normals to D at $\bar{x}$, this concludes the proof.

LEMMA 2. *A sufficient condition for $\bar{x} \in D$ and $(\bar{u}, \bar{u}^*) \in \Gamma$ to solve $(\mathscr{L}_W)$ is that $-\bar{u}^*$ be normal to D at $\bar{x}$ and $(\bar{x}, \bar{u}^*) \in \Gamma$. If f satisfies $(*)$ at $\bar{x}$, this condition is also necessary.*

Proof. Assume first that $(\bar{x}, \bar{u}^*) \in \Gamma$ and $-\bar{u}^*$ is normal to D at $\bar{x}$. Then

$$f(\bar{x}) \leqq f(u) + \langle \bar{x} - u, \bar{u}^* \rangle, \qquad \forall u \in X,$$

and

$$\langle \bar{x}, \bar{u}^* \rangle \leqq \langle x, \bar{u}^* \rangle, \qquad \forall x \in D.$$

Since we always have

$$f(u) + \langle \bar{x} - u, u^* \rangle \leqq f(\bar{x}), \qquad \forall (u, u^*) \in \Gamma,$$

it follows that

$$f(u) + \langle \bar{x} - u, u^* \rangle \leqq f(\bar{u}) + \langle \bar{x} - \bar{u}, \bar{u}^* \rangle \leqq f(\bar{u}) + \langle x - \bar{u}, \bar{u}^* \rangle$$

holds for all $(u, u^*) \in \Gamma$ and $x \in D$, that is, $(\bar{x}, \bar{u}, \bar{u}^*)$ solves $(\mathscr{L}_W)$. Now suppose $(\bar{x}, \bar{u}, \bar{u}^*)$ solves $(\mathscr{L}_W)$. The right-hand half of the saddle point inequality implies (since $f(\bar{u})$ is finite) that $-\bar{u}^*$ is normal to D at $\bar{x}$. The left-hand half yields

$$\sup_{(u,u^*) \in \Gamma} \{f(u) + \langle \bar{x} - u, u^* \rangle\} \leqq f(\bar{u}) + \langle \bar{x} - \bar{u}, \bar{u}^* \rangle.$$

If f satisfies $(*)$ at $\bar{x}$, this implies that

$$f(\bar{x}) + \langle x - \bar{x}, \bar{u}^* \rangle \leqq f(\bar{u}) + \langle x - \bar{u}, \bar{u}^* \rangle, \qquad \forall x \in X.$$

But since $(\bar{u}, \bar{u}^*) \in \Gamma$,

$$f(\bar{u}) + \langle x - \bar{u}, \bar{u}^* \rangle \leqq f(x), \qquad \forall x \in X.$$

It follows that

$$f(\bar{x}) + \langle x - \bar{x}, \bar{u}^* \rangle \leqq f(x), \qquad \forall x \in X,$$

which means $(\bar{x}, \bar{u}^*) \in \Gamma$.

Several variants and refinements of the Theorem are suggested by an examination of its proof. The constraint qualification was used only in Lemma 1 in order to guarantee the validity of the formula $\partial(f + h) = \partial f + \partial h$. Thus, the constraint qualification can be replaced by any other condition which guarantees this subdifferential formula. For the case $X = R^n$, a variant of the Theorem can therefore be derived using [7, Theorem 23.8]. This involves a slightly different constraint qualification phrased in terms of relative interiors, and it in turn allows a weakening of its hypothesis in the case that either D or f is polyhedral. In a similar way, another variant of the Theorem, valid for an arbitrary space X as above, can be derived using [3, Sect. 10e]. In these variants the condition (*) must also, of course, be reckoned with. It is a nontrivial fact, established by Brøndsted and Rockafellar [1], that f satisfies (*) everywhere when f is lower semicontinuous and X is a Banach space. A local condition sufficient (but not necessary) in order that f satisfies (*) at $\bar{x}$ is simply that $(\bar{x}, u^*) \in \Gamma$ for some $u^* \in X^*$. The complete formulation of these variants of the Theorem (and of similar variants of the result to follow) is left to the reader.

The following consequence of the Theorem generalizes White's saddle point characterization to nondifferentiable functions on spaces of arbitrary dimension.

COROLLARY. *Let $\bar{x} \in D$ and suppose f is finite and continuous at $\bar{x}$. Then $\bar{x}$ solves $(\mathscr{P})$ if and only if there exists a pair $(\bar{u}, \bar{u}^*)$ such that $(\bar{x}, \bar{u}, \bar{u}^*)$ solves $(\mathscr{L}_W)$.*

Proof. Observe that

$$f(x) \leqq \sup_{(u,u^*)\in\Gamma} \{f(u) + \langle x - u, u^* \rangle\} \leqq f(x)$$

holds whenever x lies in the projection of Γ onto X, i.e., whenever $\varnothing \neq \partial f(x)$. Now a sufficient condition for $\varnothing \neq \partial f(x)$ is that f be finite and continuous at x [3, Sect. 10c]. Hence the hypothesis implies that f satisfies (*) at $\bar{x}$. Since the hypothesis implies also that $(\mathscr{P})$ satisfies the constraint qualification, the conclusion is now immediate from the theorem.

We turn now to question (ii) of the Introduction. Using a technique well known in duality theory, we regard $(\mathscr{L}_W)$ as a "Lagrangian" problem and split it into two related problems, a "primal" problem $(\mathscr{P}_W)$

$$\min_{x \in D} \psi(x), \qquad \text{where} \qquad \psi(x) = \sup_{(u, u^*) \in \Gamma} J(x, u, u^*), \tag{$\mathscr{P}_W$}$$

and a "dual" problem

$$\max_{(u, u^*) \in \Gamma} \varphi(u, u^*), \qquad \text{where} \qquad \varphi(u, u^*) = \inf_{x \in D} J(x, u, u^*). \tag{$\mathscr{D}_W$}$$

Notice that as long as f satisfies (*), $\psi = f$ and hence $(\mathscr{P}_W)$ coincides with $(\mathscr{P})$. Thus in particular, $(\mathscr{P}_W)$ coincides with $(\mathscr{P})$ when f is finite and continuous on X.

If $(\bar{x}, \bar{u}, \bar{u}^*)$ solves $(\mathscr{L}_W)$, it follows trivially that $\bar{x}$ solves $(\mathscr{P}_W)$ and $(\bar{u}, \bar{u}^*)$ solves $(\mathscr{D}_W)$. By the theorem, therefore, solvability of $(\mathscr{D}_W)$ is a necessary condition for solvability of $(\mathscr{P})$, provided $(\mathscr{P})$ satisfies the constraint qualification.

One can inquire also into the relationship between the optimal values in $(\mathscr{P})$ and $(\mathscr{D}_W)$. One always has the inequalities

$$\sup_{\Gamma} \inf_{D} J \leqq \inf_{D} \sup_{\Gamma} J \leqq \inf_{D} f,$$

i.e., $\mathrm{val}(\mathscr{D}_W) \leqq \mathrm{val}(\mathscr{P}_W) \leqq \mathrm{val}(\mathscr{P})$. The left-hand inequality is well known (e.g., [7, Lemma 36.1]) and easy to verify, while the right-hand inequality is immediate from the trivial inequality $\psi \leqq f$. Thus, the so-called "weak duality theorem" holds between $(\mathscr{D}_W)$ and $(\mathscr{P})$.

For computational as well as theoretical reasons, it is of great interest to know conditions which guarantee that actually $\mathrm{val}(\mathscr{D}_W) = \mathrm{val}(\mathscr{P})$, i.e., that there be no "duality gap." By the Theorem, a trivial such condition is that $(\mathscr{P})$ be solvable and satisfy the constraint qualification. One may still wish to compute $\mathrm{val}(\mathscr{P})$ by means of $\mathrm{val}(\mathscr{D}_W)$, however, even when it cannot be ascertained a priori that $(\mathscr{P})$ is solvable. A simple example illustrates how bad things can be taking this approach. Consider the problem $(\mathscr{P})$ determined by the elements $D = X = R$ and $f(x) = e^x$. Except for possessing no solution, this $(\mathscr{P})$ is as nice as one could wish; yet $\mathrm{val}(\mathscr{P}) = 0$ and $\mathrm{val}(\mathscr{D}_W) = -\infty$. In the next section, another example of $(\mathscr{P})$ is given in which f is continuously differentiable on $X = R^2$, D is a line (or closed half-space), $\mathrm{val}(\mathscr{P}) = -1$, but $\mathrm{val}(\mathscr{D}_W) = -3/2$.

Another important point to note is that the duality framework given by problems $(\mathscr{P}_W)$, $(\mathscr{D}_W)$, and $(\mathscr{L}_W)$ does not (in any obvious manner, at least) reflect further structure which may be present in the constraint set D. For example, if D of the form $\bigcap_{i=1}^m \{x \in X \mid f_i(x) \leqq 0\}$ for some proper convex functions $f_1, \ldots, f_m$, then one would like the duality formulation to take the

f_i's into account somehow and (hopefully) to include the usual theory of Lagrange multipliers.

A third point to notice in connection with this trio of problems is that they may not even exist in the general case, in the sense that the set Γ may be empty! Indeed, Brøndsted and Rockafellar [1], by using an example of Klee, have exhibited a lower semicontinuous f on a reflexive Fréchet space X with the property that $\Gamma = \varnothing$. On the other hand, they have shown that when f is lower semicontinuous and X is a Banach space, $\Gamma \neq \varnothing$ and moreover f satisfies (*) everywhere. Another case in which $\Gamma \neq \varnothing$ and f satisfies (*) everywhere, is that in which f is finite and continuous on X (cf. proof of Corollary).

Finally, as will be discussed in the next section, the maximization problem $(\mathscr{D}_W)$ is in general nonconcave, even when f is differentiable.

The four points just mentioned would seem to constitute significant obstacles to basing a general duality theory for convex programming either on the function H or on its generalization J discussed above. In the next section we shall relate the foregoing approach to an existing duality theory not possessing these drawbacks.

3. Conjugate Duality

The basic idea leading to White's choice of H as a Lagrangian function (and to the generalization J discussed above) is the identity (*), i.e., the representation of a convex function as the envelope of certain hyperplanes. This same idea can still be used, albeit in a somewhat different way, to develop a comprehensive duality theory for convex programming problems. This has already been done, in fact, and it is rendered possible essentially by making fuller use of the dual space X^*. The key tool is Fenchel's conjugacy correspondence among convex functions. In this section we review briefly an important special case of the resulting duality framework and then relate it to the trio of problems $(\mathscr{P}_W)$, $(\mathscr{D}_W)$, and $(\mathscr{L}_W)$.

Let X, X^*, f, and the bilinear form $\langle \cdot, \cdot \rangle$ be as in Section 2, and assume additionally that the given topology on X is locally convex. The *conjugate* of f is the function f^* on X^* given by

$$f^*(x^*) = \sup_{x \in X} \{\langle x, x^* \rangle - f(x)\}.$$

The function f^* is proper convex and lower semicontinuous. Clearly the condition $x^* \in \partial f(x)$ is equivalent to $f(x) + f^*(x^*) = \langle x, x^* \rangle$, and these conditions imply $x \in \partial f^*(x^*)$. (We regard X as a subspace of its bidual X^{**}.) One can repeat the conjugacy operation, obtaining from f^* a function f^{**}

on X^{**} having the same properties. It is known that f^{**} agrees with f on X when f is lower semicontinuous, so that $x \in \partial f^*(x^*)$ implies $x^* \in \partial f(x)$ in that case. For a proper concave function g (i.e., $-g$ is proper convex), one makes parallel definitions, the general rule being to interchange supremum with infimum, $+\infty$ with $-\infty$, and $\geqq$ with $\leqq$. (NB: This results in $g^*(x^*) = -(-g)^*(-x^*)$.) Henceforth, let g be a proper concave function on X and use the notation

$$C = \{x \in X \mid f(x) < +\infty\}, \qquad C^* = \{x^* \in X^* \mid f^*(x^*) < +\infty\},$$

$$D = \{x \in X \mid g(x) > -\infty\}, \qquad D^* = \{x^* \in X^* \mid g^*(x^*) > -\infty\}.$$

In 1951, W. Fenchel [2] considered the problem

$$\min_X \{f - g\} \tag{$\mathscr{P}_F$}$$

for the case $X = R^n$ and proved a fundamental duality theorem relating it to the problem

$$\max_{X^*} \{g^* - f^*\}. \tag{$\mathscr{D}_F$}$$

Fenchel's theorem states that, if the relative interiors of C and D have a point in common, $\mathrm{val}(\mathscr{P}_F) = \mathrm{val}(\mathscr{D}_F)$ and moreover $(\mathscr{D}_F)$ is solvable. There is an obvious dual assertion involving solvability of $(\mathscr{P}_F)$ in the event f and g are, respectively, lower and upper semicontinuous, due to the symmetric nature of the conjugacy correspondence in that case. For the function L on $C \times D^*$ given by

$$L(x, u^*) = f(x) + g^*(u^*) - \langle x, u^* \rangle,$$

one can compute directly that $(\mathscr{D}_F)$ coincides with the problem

$$\max_{u^* \in D^*} \{\inf_{x \in C} L(x, u^*)\},$$

while $(\mathscr{P}_F)$ coincides with the problem

$$\min_{x \in C} \{\sup_{u^* \in D^*} L(x, u^*)\}$$

when g is upper semicontinuous. Thus, intimately related to $(\mathscr{P}_F)$ and $(\mathscr{D}_F)$ is the saddle point problem

$$\operatorname*{minimax}_{x \in C \ u^* \in D^*} L(x, u^*). \tag{$\mathscr{L}_F$}$$

We shall not take the space here to restate the many existing results concerning the trio $(\mathscr{P}_F)$, $(\mathscr{D}_F)$, and $(\mathscr{L}_F)$ and its generalizations, but instead just refer the reader to a selection of the extensive literature [2, 4, 5, 7, 8].

From now on we consider the situation in which g is specialized to be the function which is 0 on D and $-\infty$ on $X \backslash D$. Then $(\mathscr{P}_F)$ is precisely $(\mathscr{P})$, and $(\mathscr{D}_F)$ is clearly the same as

$$\max_{D^* \cap C^*} \{g^* - f^*\}, \tag{$\mathscr{D}_F'$}$$

where for our choice of g we have

$$g^*(u^*) = \inf_{x \in D} \langle x, u^* \rangle.$$

Fenchel's dual $(\mathscr{D}_F)$ is a concave maximization problem. On the other hand, observe that the identity

$$L(x, u^*) = \inf_{u \in D} J(u, x, u^*) = \varphi(x, u^*)$$

holds as long as $x \in C$, while if actually $(x, u^*) \in \Gamma$, this quantity equals $g^*(u^*) - f^*(u^*)$. Thus, although L is convex-concave on $C \times X^*$, the function φ is actually concave on $\Gamma \subset C \times X^*$. However, since Γ and its projection onto X^* are in general not convex sets, $(\mathscr{D}_W)$ is in general a non-concave problem (cf. Example 2 below). Notice, incidentally, that when f is differentiable on X one can interpret φ on the X-space alone as

$$\varphi(u, \nabla f(u)) = g^*(\nabla f(u)) - f^*(\nabla f(u)),$$

so that $(\mathscr{D}_W)$ is the same as

$$\max_{u \in X} \{g^*(\nabla f(u)) - f^*(\nabla f(u))\}.$$

Even in this formulation $(\mathscr{D}_W)$ is not concave, though, since the mapping $u \to g^*(\nabla f(u)) - f^*(\nabla f(u))$ need not be concave.

Fenchel's trio of problems, unlike White's, is well defined even in the general case in which Γ may be empty. Also, within Fenchel's framework one can explicitly deal with additional structure which may be present in $(\mathscr{P})$ via the set D. For a derivation of the theory of Lagrange multipliers in this context, see [4, Sect. 6]. We conclude the paper with a comparison of the duality gaps associated with Fenchel's and White's dual problems.

Since $(u, u^*) \in \Gamma$ is equivalent to $f(u) - \langle u, u^* \rangle = -f^*(u^*)$, we can express $(\mathscr{D}_W)$ equivalently as

$$\max_{D^* \cap \tilde{C}^*} \{g^* - f^*\}, \tag{$\mathscr{D}_W'$}$$

where

$$\tilde{C}^* = \{u^* \in X^* \mid (u, u^*) \in \Gamma \text{ for some } u \in X\}.$$

Since $\tilde{C}^* \subset C^*$, a comparison of $(\mathscr{D}_W')$ with $(\mathscr{D}_F')$ implies immediately that

$$\operatorname{val}(\mathscr{D}_W) \leqq \operatorname{val}(\mathscr{D}_F) \leqq \operatorname{val}(\mathscr{P}).$$

Thus, the Fenchel dual always yields at least as small a duality gap as does White's dual. It can happen, moreover, that the White duality gap is positive or infinite at the same time that the Fenchel duality gap is zero (and, moreover, the Fenchel dual $(\mathscr{D}_F)$ is solvable). This behavior can occur even with continuously differentiable functions on spaces of finite dimension, as shown by the following two examples.

EXAMPLE 1 (from Sect. 2). For $X = D = R$ and $f(x) = e^x$ one has

$$\sup_{D^* \cap \tilde{C}^*} \{g^* - f^*\} = -\infty < 0 = \max_{D^* \cap C^*} \{g^* - f^*\} = \inf_D f.$$

The right-hand equality follows from the version of Fenchel's Duality Theorem stated above, while the value 0 follows by inspection of $(\mathscr{P})$. That the left-hand value is $-\infty$ can be verified by direct computation from the original formulation of $(\mathscr{D}_W)$. It is instructive, however, to evaluate the expression directly in X^* via g^*, f^*, and $\tilde{C}^*$. Clearly g^* is 0 at the origin and $-\infty$ elsewhere, so that $D^* = \{0\}$. As an exercise one can compute that

$$f^*(u^*) = \begin{cases} u^* \ln u^* - u^*, & u^* > 0 \\ 0, & u^* = 0 \\ +\infty, & u^* < 0, \end{cases}$$

so that $C^* = [0, +\infty)$. Because f is lower semicontinuous, $(u, u^*) \in \Gamma$ if and only if $u \in \partial f^*(u^*)$, i.e.,

$$f^*(x^*) \geqq f^*(u^*) + \langle u, x^* - u^* \rangle, \qquad \forall x^* \in X^*.$$

It follows by inspection of the graph of f^* that $\tilde{C}^* = (0, +\infty)$. Hence $D^* \cap \tilde{C}^* = \varnothing$, which implies (by the conventions $\sup_\varnothing = -\infty$ and $\inf_\varnothing = +\infty$) that

$$\sup_{D^* \cap \tilde{C}^*} \{g^* - f^*\} = -\infty.$$

EXAMPLE 2. There exists a continuously differentiable f on $X = R^2$ such that for $D = \{(\xi_1, \xi_2) \mid \xi_2 = 0\}$ one has

$$\sup_{D^* \cap \tilde{C}^*} \{g^* - f^*\} = -\tfrac{3}{2} < -1 = \max_{D^* \cap C^*} \{g^* - f^*\} = \inf_D f.$$

Unlike in the previous example, here the set $D^* \cap \tilde{C}^*$ will be nonempty. Observe first that, for D as above, g^* is 0 along the second coordinate axis and

$-\infty$ elsewhere, so that $D^* = \{(\eta_1, \eta_2) \mid \eta_1 = 0\}$. We now construct indirectly a function f having the asserted properties. Our method is to construct a function we call f^* and then take f to be $(f^*)^*$. Consider the function

$$h(\eta_1, \eta_2) = \max\{h_1(\eta_1), h_2(\eta_2)\},$$

where

$$h_1(\eta_1) = \begin{cases} 1 - (\eta_1)^{\frac{1}{2}} & \text{if } \eta_1 \in [0, 1] \\ +\infty & \text{otherwise} \end{cases} \qquad h_2(\eta_2) = \begin{cases} |\eta_2| & \text{if } \eta_2 \in [-1, 1] \\ +\infty & \text{otherwise.} \end{cases}$$

This h is proper convex and lower semicontinuous, and it can be seen that $(0, 1)$ and $(0, -1)$ are the only points of D^* at which h is subdifferentiable (cf. [7, p. 218]). Put $f^* = h + w$, where

$$w(\eta_1, \eta_2) = \tfrac{1}{2}\|(\eta_1, \eta_2)\|^2 = \tfrac{1}{2}(\eta_1{}^2 + \eta_2{}^2).$$

Then f^* is proper convex and lower semicontinuous [7, Theorem 9.3], $C^* = [0, 1] \times [-1, 1]$, and moreover $(0, 1)$ and $(0, -1)$ are the only points of D^* at which f^* is subdifferentiable [7, Theorem 23.8]. Hence

$$D^* \cap C^* = \{0\} \times [-1, 1] \qquad \text{and} \qquad D^* \cap \tilde{C}^* = \{0\} \times \{-1, 1\},$$

whence

$$\sup_{D^* \cap \tilde{C}^*} \{g^* - f^*\} = -\tfrac{3}{2} < -1 = \sup_{D^* \cap C^*} \{g^* - f^*\}.$$

Now put $f = (f^*)^*$. (The one-to-one symmetric nature of the conjugacy correspondence ensures that the conjugate of $(f^*)^*$ is indeed the function $h + w$ constructed above.) Using the fact that w^* has the same formula as w, we conclude from [7, Corollary 26.3.2] that f is differentiable on int $C \neq \varnothing$. But since C^* is bounded, $C = X$ (e.g. [7, Corollary 13.3.1 or Theorem 13.4]). Hence f is continuously differentiable everywhere on X [7, Corollary 25.5.1]. Finally, $\text{val}(\mathscr{D}_F) = \text{val}(\mathscr{P})$ with $(\mathscr{D}_F)$ solvable follows by the previously mentioned theorem of Fenchel.

We conclude by giving conditions under which the White duality gap coincides with the Fenchel duality gap.

LEMMA 3. *One has* $\text{val}(\mathscr{D}_W) = \text{val}(\mathscr{D}_F)$ *under either of the following circumstances:* (a) $X = R^n$ *and the relative interiors of* D^* *and* C^* *contain a point in common;* (b) X *is a reflexive Banach space,* D *is closed, and for some* $\alpha > \inf_X f$ *the set* $\{x \in X \mid f(x) \leqq \alpha\}$ *is weakly compact.*

Proof. The assertion is that

$$\sup_{D^* \cap \tilde{C}^*} \{g^* - f^*\} \geqq \sup_{D^* \cap C^*} \{g^* - f^*\},$$

or equivalently that

$$\inf_{D^* \cap \tilde{C}^*} h \leqq \inf_{X^*} h,$$

where $h = f^* - g^*$. Consider first case (a). Clearly h is proper convex with $\operatorname{dom} h = \{x^* \in X^* \mid h(x^*) < +\infty\} = D^* \cap C^*$. By [7, Theorem 23.4], $\tilde{C}^* \supset \operatorname{ri} C^*$ where "ri" denotes the relative interior operation. The hypothesis implies by [7, Theorem 6.5] that $\operatorname{ri}(D^* \cap C^*) = \operatorname{ri} D^* \cap \operatorname{ri} C^*$, so that $D^* \cap \tilde{C}^* \supset \operatorname{ri}(\operatorname{dom} h)$. Hence

$$\inf_{D^* \cap \tilde{C}^*} h \leqq \inf_{\operatorname{ri}(\operatorname{dom} h)} h = \inf_{X^*} h,$$

where the equality is a consequence of [7, Theorems 7.5 and 6.1]. Now consider case (b), and let k be the function which is 0 on $-D$ and $+\infty$ on $X \backslash D$. By [3, Sect. 4d] the function $f \,\square\, k$, i.e., the infimal convolution of f with k, is proper convex and lower semicontinuous. Hence, so is its conjugate $(f \,\square\, k)^* = f^* - g^* = h$. It follows from [1, Theorem 2] that

$$\inf_{X^*} h = \inf_{\operatorname{dom} \partial h} h,$$

where

$$\operatorname{dom} \partial h = \{u^* \in X^* \mid u \in \partial h(u^*) \text{ for some } u \in X\}.$$

Now by [3, Sect. 8f], f^* is finite and continuous at the origin with respect to some admissible topology on X^*, so that

$$\partial h(u^*) = \partial f^*(u^*) - \partial g^*(u^*), \qquad \forall u^* \in X^*$$

by [3, Sect. 10d]. Hence $\operatorname{dom} \partial h \subset \tilde{C}^* \cap D^*$, so that the infimum of h over $\tilde{C}^* \cap D^*$ does not exceed that of h over $\operatorname{dom} \partial h$. This concludes the proof.

Unfortunately, the conditions in case (b) of the lemma are so strong as to imply also that $\operatorname{val}(\mathscr{D}_F) = \operatorname{val}(\mathscr{P}_F)$ and that in fact $(\mathscr{P}_F)$ is solvable (see [4, Corollary 2] or [5, Corollary to Theorem 1]). If D is closed and f is lower semicontinuous, the conditions in case (a) also imply these additional conclusions (see [7, Theorem 31.1]).

References

1. A. Brøndsted and R. T. Rockafellar, On the subdifferentiability of convex functions, *Proc. Amer. Math. Soc.* **16** (1965), 605–611.
2. W. Fenchel, Convex Cones, Sets and Functions, mimeographed lecture notes, Princeton University, Spring 1951.
3. J.-J. Moreau, Fonctionelles Convexes, mimeographed lecture notes, Séminaire sur les Équations aux Dérivées Partielles, Collège de France, 1966–1967.

4. R. T. ROCKAFELLAR, Extension of Fenchel's duality theorem for convex functions, *Duke Math. J.* **33** (1966), 81–89.
5. R. T. ROCKAFELLAR, Duality and stability in extremum problems involving convex functions, *Pacific J. Math.* **21** (1967), 167–187.
6. R. T. ROCKAFELLAR, Convex functions, monotone operators and variational inequalities, *in* "Theory and Applications of Monotone Operators" (A. Ghizzetti, Ed.), 35–65, Oderisi, Gubbio, Italy, 1968.
7. R. T. ROCKAFELLAR, "Convex Analysis," Princeton University Press, Princeton, NJ, 1970.
8. J. STOER AND C. WITZGALL, "Convexity and Optimization in Finite Dimensions I," Springer-Verlag, New York, 1970.
9. D. J. WHITE, Envelope programming and a minimax theorem, *J. Math. Anal. Appl.* **40** (1972), 1–11.